Response to Referee 2

for "Representation learning with unconditional denoising diffusion models for dynamical systems"

Finn, T. S., Disson, L., Farchi, A., Bocquet, M., and Durand, C.

24th May 2024

RC: Reviewer Comment, AR: Author Response

RC: **This is a very interesting and novel study on the use of denoising diffusion model for representation learning. The manuscript is well written and describes very nicely the context, how these approaches (rooted in image applications) can be adapted to geosciences, and illustrates two distinct relevant applications, surrogate modelling and ensemble generations, that are both extremely important in high dimensional settings.**
I think the manuscript can be accepted almost as it is, but I have a few minor comments I would encourage the Authors to look at.

AR: We thank the second referee for the constructive feedback on our manuscript, especially with the remarks related to the theory of dynamical systems. In the following, we will discuss the raised comments and indicate what we will change in the revised manuscript.

RC: **While there are little spaces for doubts, I would strongly suggest the Authors to specify that their approach applies to ergodic chaotic dynamics for which an invariant distribution exists that describe the state distribution on the system's attractor. An obvious counterexample would be a stable system having an equilibrium point (or a limit cycle) as attractor.**

AR: Thank you for the suggestion to sharpen the focus of the manuscript around ergodic chaotic dynamics. To sharpen the introduction and take this suggestion into account, we will add to the second paragraph in the introduction "This generation process is expected to be successful if there is an invariant state distribution on the system's attractor, which exists for ergodic chaotic dynamics."

RC: **When mentioning the Schrodinger Bridge (page 2), you may want to refer to Reich S. 2019 (doi:10.1017/S0962492919000011) as an exemplar study of the same analogy but in the area of data assimilation.**

AR: Thank you for pointing us to the seminal work of S. Reich on the connection between the Schrödinger Bridge and data assimilation. Originally, we left this work out as our study is focused on machine learning. However, since we apply diffusion models on a data assimilation problem, we will add this work to the introduction.

RC: Line 27. "..dynamical systemS ..."

AR: Thank you for spotting this inconsistency, we will use the plural of dynamical systems as proposed.

RC: In the caption of Fig1b, use (left/right) to point the reader.

AR: Thank you for the suggestion, we will add this to the caption.

RC: Line 44. I think you should always order references chronologically.

AR: Thank you for spotting the chronological inconcistency here, we will swap the citations.

RC: Line 53–59. While I understand and I like the Authors narrative and choice of references. Nevertheless, and particularly for the readers of NPG, it would be appropriate to also mention the large bulk of work on the generation of ensemble members based on dynamical systems's theory and data assimilation. A good recent reference is 10.1029/2021MS002828

AR: Thank you for the suggestion of adding references about the generation of ensemble members based on the properties of the dynamical system. To take these works into account, we will add to the paragraph *"Another method to generate an ensemble for data assimilation would be to make use of the knowledge about the system's error propagation, in form of singular vectors (Molteni et al., 1996) or bred vectors (Toth and Kalnay, 1993), as they are similarly used to initialize ensemble weather forecasts (Buizza et al. 2005) or subseasonal forecasts (Demaeyer et al., 2022)".*

RC: I am a bit of an inconvenience with the use of the term "latent". On the one side I agree with a comment from the other Reviewer. On the other I do also see in line 100 that you state z=x which makes one deduce the latent and actual state have the same dimension. Finally, while it is true that latent variables are defined in relation to their indirect (often hidden) relation with the observables quantities, with no reference to their number (or space dimension), in many practical applications the latent space is assumed/defined/used as being of smaller dimension.

AR: We agree with both referees about the confusion with the latent space of diffusion models, especially if they are used for representation learning where the features of the trained NN span another space. The latent space of diffusion models usually describes the space of noised data, while the other is rather a *feature space*. The noised data space is called latent space as diffusion models can be seen as hierarchical variational autoencoder (Luo, 2022), where one can also learn a mapping from data space into this latent space with a possibly reduced dimensionality (e.g., Rombach et al., 2022). Diffusion models that act in data space can be seen as special version of these variational diffusion models, where the mapping between spaces is the identity function.

RC: Line 115. I would add "…. prior distribution FOR THE DENOISING PROCESS."

AR: Thank you for the suggestion and we will add *"for the diffusion process"* to make the writing clearer.

RC: Equations (8). Wouldn't be better to (re)state clearly that we do not have access to x in practice?

AR: Thank you for the suggestion and we will add *"Note, during generation, the state* $\mathbf{x}$ *is unknown and we have to approximate Eq. (8b) to generate data, as we discuss later."* to clarify the use of the Equations 8 for generation.

RC: Line 145. Is that because they do not depend on x?

AR: As the neural network just approximates the denoised state by Tweedie's formula, we have defined in Eq. (10), that the covariances between the analytical denoising step and the approximated step match, other definition are possible though (e.g., Ho et al., 2020, Nichol and Dhariwal, 2021). With the definition of the diffusion process, we can then specify the Kullback-Leibler divergence with different quantities, e.g., if we would directly predict the state $\mathbf{x}_\theta(\mathbf{z}_\tau, \tau)$ or the noise $\boldsymbol{\epsilon}_\theta(\mathbf{z}_\tau, \tau)$. The simplification of the loss can be traced in Ho et al. (2020) or in Luo (2022).

RC: Line 153. I think "Equation" must be written at the beginning of the sentence.

AR: Thank you for spotting this, yes, we agree that *Equation* should be written at the beginning of a sentence and we will correct the sentence accordingly.

RC: Line 176. Instead of "normally" I would suggest "most of the times".

AR: Thank you for the nice suggestion, *"most of the times"* sounds indeed better and we will change the words accordingly.